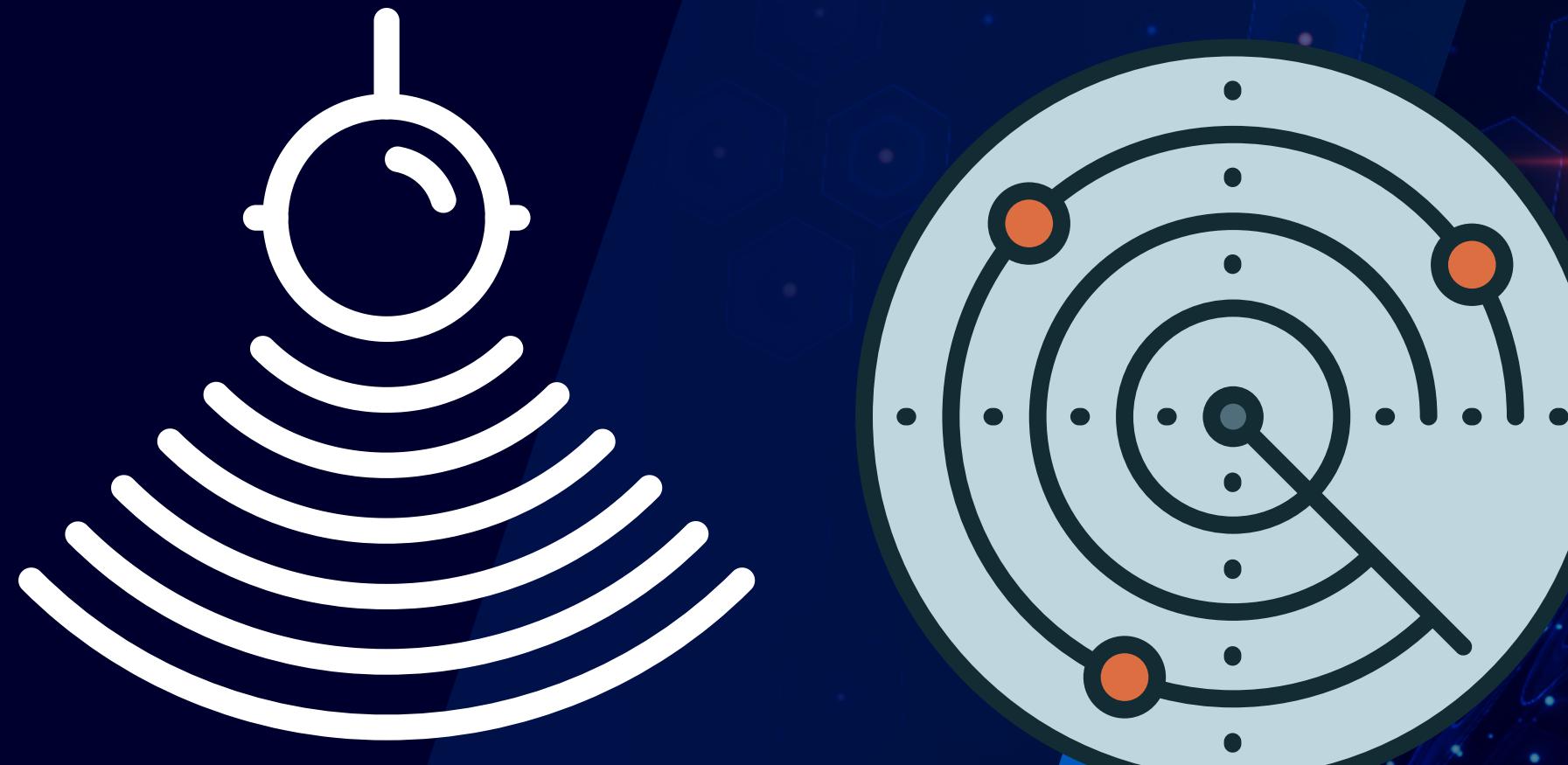




ROCK OR MINE: A SONAR CLASSIFICATION CHALLENGE



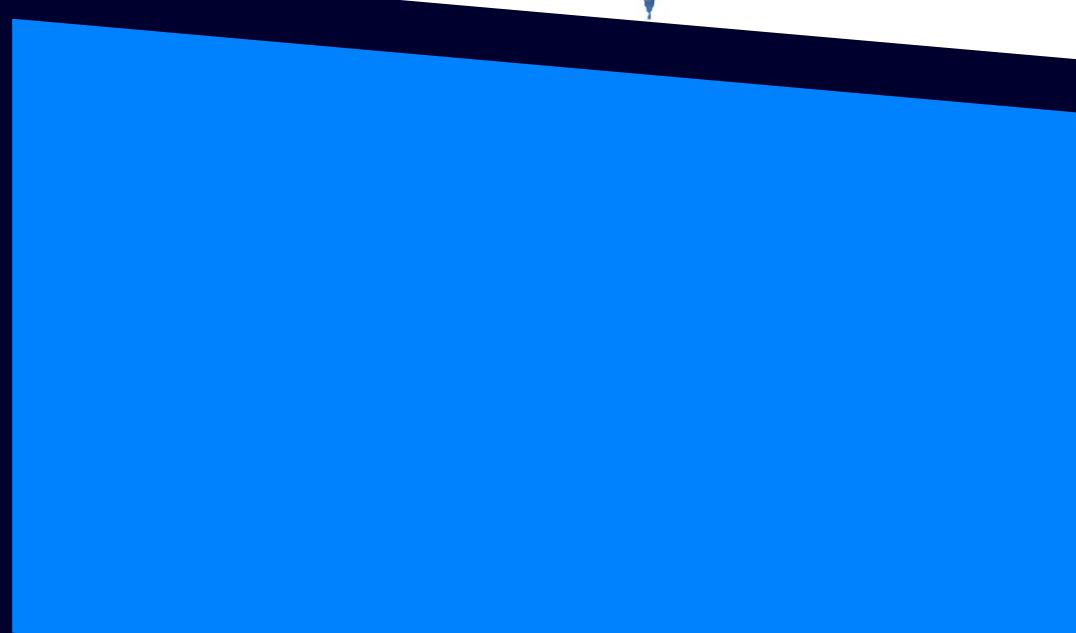
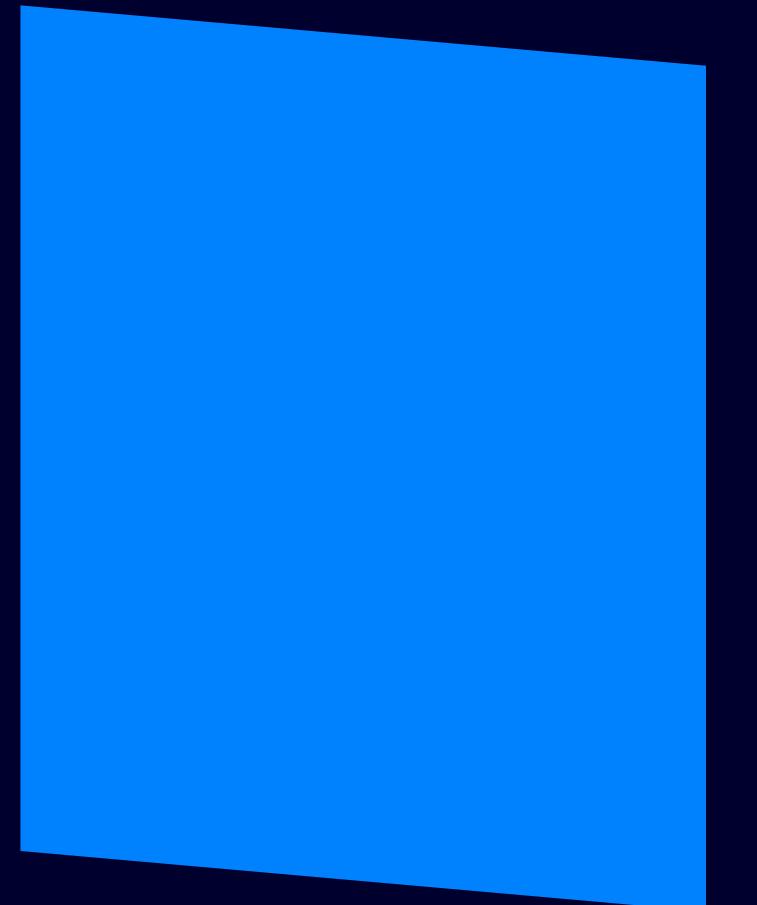
PROJECT OVERVIEW

- *The Sonar Dataset is a classification problem that aims to classify sonar signals into two classes - either a rock or a mine. The dataset consists of 208 observations, each with 60 features.*
- *The goal of this project is to train a machine learning model to accurately classify the sonar signals and evaluate the performance of two popular algorithms - KNN and Logistic Regression.*

DATASET DESCRIPTION

- *The Sonar Dataset was obtained from the UCI Machine Learning Repository. The dataset consists of 208 observations, each with 60 features. The dataset has been preprocessed and normalized between -1 and 1.*

METHODOLOGY



DATA EXPLORATION AND VISUALIZATION

- We started by exploring the dataset using Python libraries such as Pandas, NumPy, and Matplotlib.
- We created various visualizations, such as histograms, box plots, and scatter plots, to understand the distribution and correlation of features.



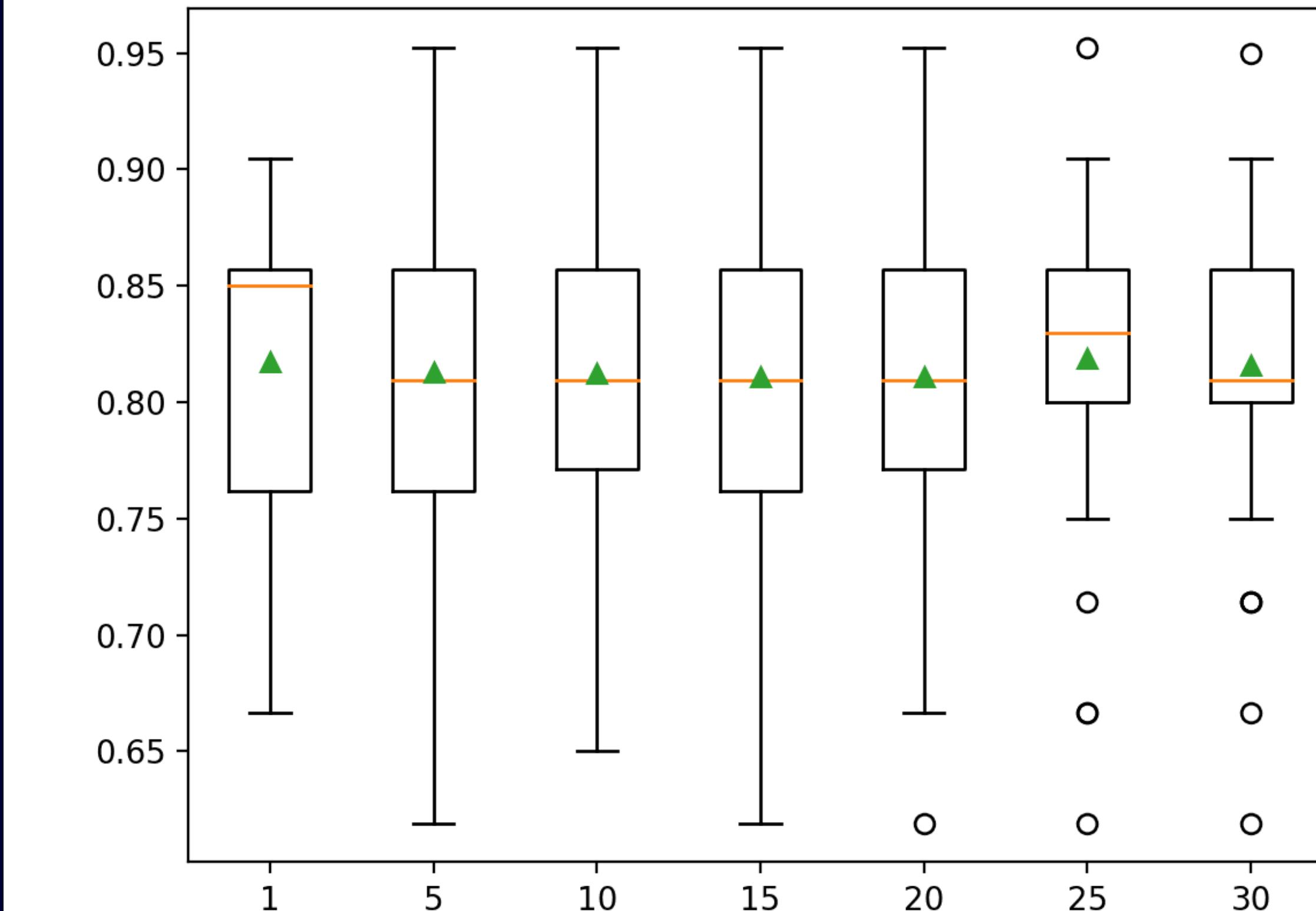
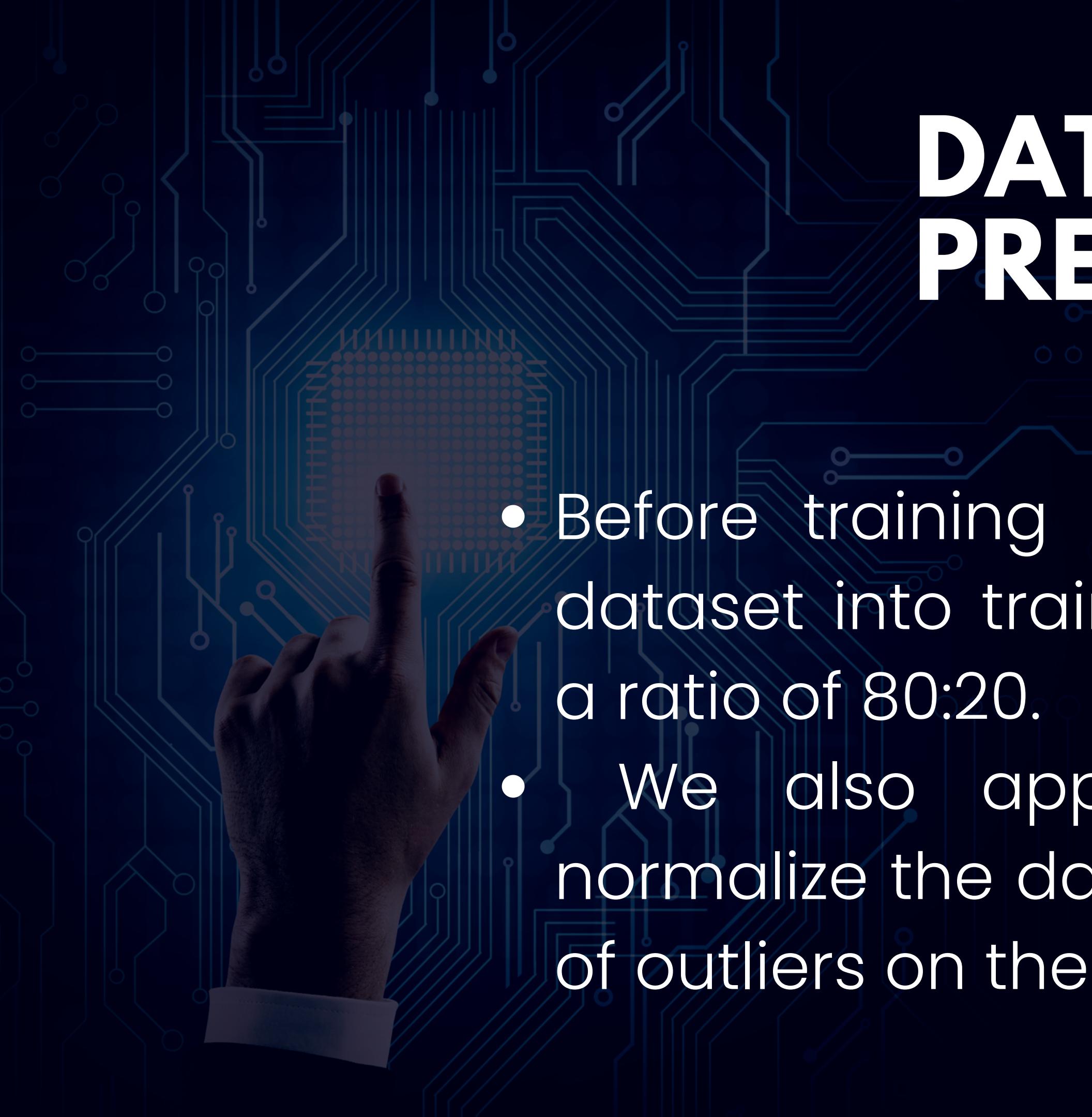


Figure 1: Boxplot of the Sonar Dataset



DATA PREPROCESSING

- Before training the models, we split the dataset into training and testing sets with a ratio of 80:20.
- We also applied feature scaling to normalize the data and reduce the impact of outliers on the models.

Model Training

We trained two popular classification algorithms - **KNN** and **Logistic Regression** - on the Sonar Dataset.



KNN MODEL

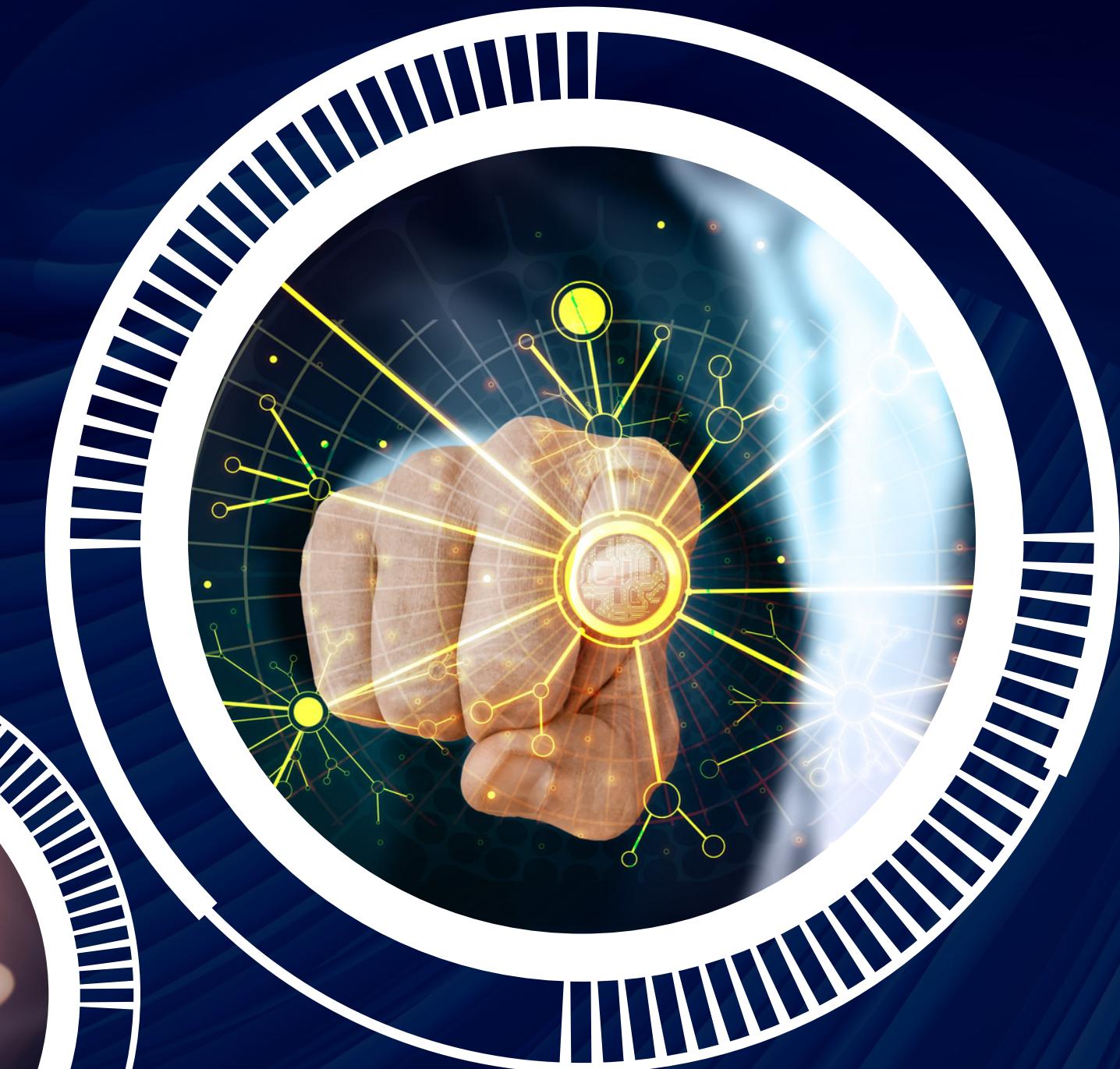
- We trained a KNN model with $k=3$ using the training data.
- We used the Euclidean distance metric to calculate the distance between instances.
- We evaluated the performance of the model using accuracy, precision, recall, F1 score, and the confusion matrix.

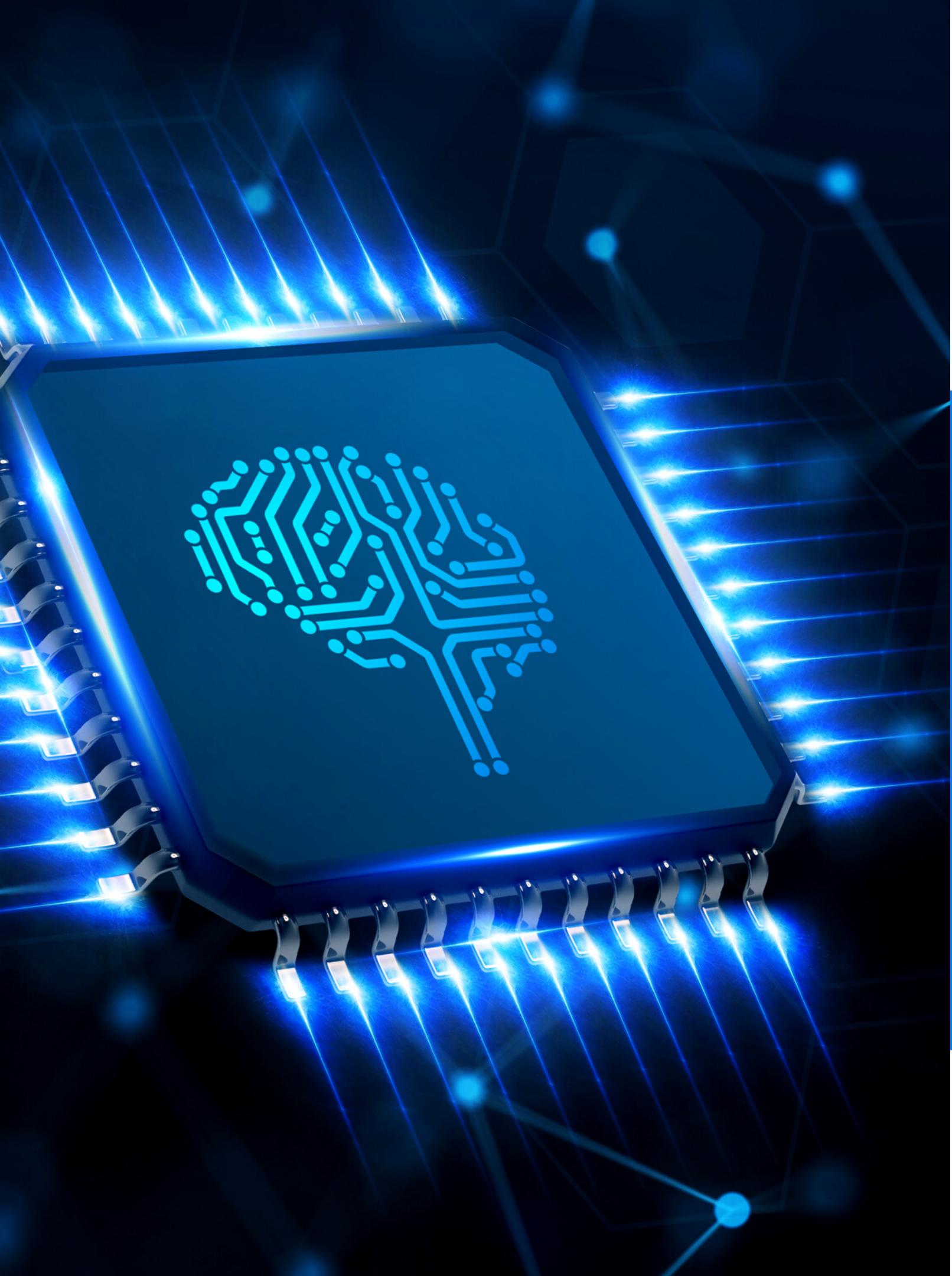
LOGISTIC REGRESSION MODEL



- We also trained a *Logistic Regression* model using the training data.
- We used *L2 regularization* and optimized the model using the *lbfgs* solver.
- We evaluated the performance of the model using the same metrics as the *KNN* model.

RESULTS





LOGISTIC REGRESSION MODEL RESULTS

- *The Logistic Regression model achieved an accuracy of 83.7% on the test data.*
- *The precision, recall, and F1 score were all above 76%, indicating good performance.*
- *The confusion matrix shows that the model was able to correctly classify 19 out of 21 mines and 10 out of 13 rocks.*

accuracy on training data

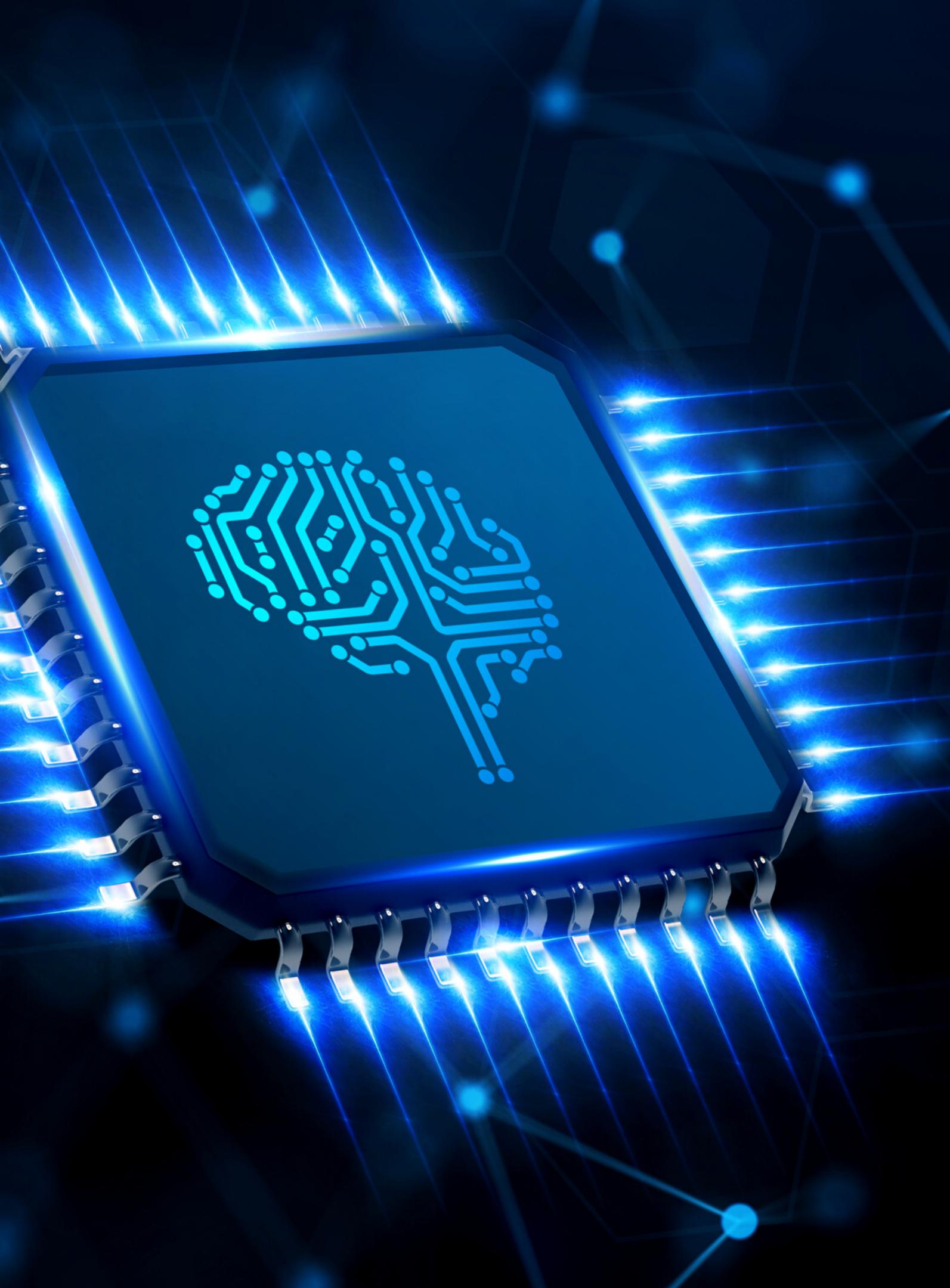


83.2%

accuracy on test data:



76%



KNN MODEL RESULTS

- *The KNN model achieved an accuracy of 89% on the test data.*
- *The precision, recall, and F1 score were all above 90%, indicating good performance.*
- *The confusion matrix shows that the model was able to correctly classify 19 out of 21 mines and 11 out of 13 rocks.*

KNN model accuracy of training data is

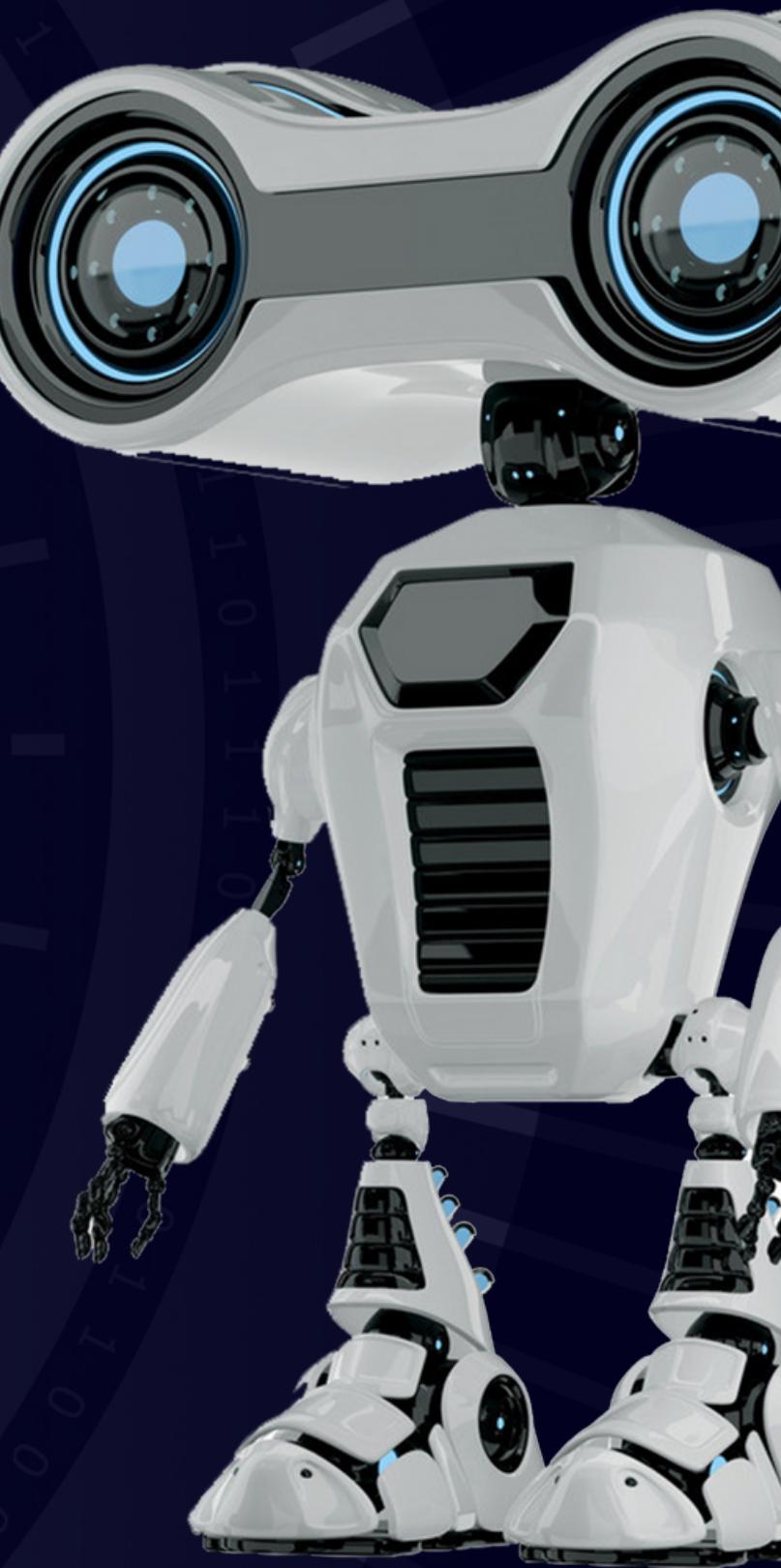


KNN model accuracy of test data is



CONCLUSION

- In conclusion, we were able to train two popular classification algorithms - **KNN** and **Logistic Regression** - on the Sonar Dataset.
- Both models achieved good performance, with the KNN model slightly outperforming the Logistic Regression model.
- Overall, the results demonstrate that machine learning algorithms can effectively classify sonar signals, which has important applications in underwater navigation and surveillance.





THANK YOU
FOR YOUR ATTENTION